

River Basin Model: Proposed Data Services for Columbia River Temperature (TMDL)

Ross Yearsley

EPA Region 10, Office of Environmental Assessment

January 28, 2003

Executive Summary

The River Model Data Service is a 3-tier application that will integrate into current TMDL processes. The River Data System will provide a central data store that will insure integrity of original data and support efficient querying, thereby improving the current system, which uses flat files for data storage. Initially the River Basin Model will provide the following functionality:

1. Access database containing flow, temperature and elevation data necessary to run simulations.
2. Graphical User Interface that will allow the user to create temperature, flow and elevation data files which can then be processed for incomplete/erroneous data by existing utilities.
3. If time permits, functionality will be added to allow users to specify a particular scenario. The River Basin Model will then automatically generate the initial setup file that contains the ordering for the simulation.

I will provide the database design and implementation, application logic necessary to query the database and a graphical user interface to act as a front end. I will input a limited amount of data that will be used to test the system.

TMDL Overview

The Columbia River TMDL is a system that can effectively predict water temperatures in the Columbia and other rivers under a variety of scenarios. At the heart of the TMDL is a computer program that requires large amounts of preprocessed data in order to run. While the TMDL was originally developed to simulate Columbia River temperatures, the simulation process can be applied to any river or series of rivers for which there are sufficient data.

In order to successfully predict temperatures for a given river, the TMDL requires the following inputs:

- Weather data (e.g. air temperature, cloud cover)
- Flow and water temperature data for tributaries at confluence points
- Water elevation data at dams
- River topology (unlike flow, elevation, temperature and weather, topology is more or less static and does not vary with time)

Typically, a simulation is run in time steps of one day over a period of multiple days, months or years. For example, a given scenario might simulate temperatures in the Columbia from January 15, 2000 to September 30, 2001. At each time step the program expects appropriate data.

River System Model

In the TMDL, a river or river system is represented as a directed acyclic graph. In order to run a simulation, a scenario is created and an ordering is constructed. Presently, the ordering process is done manually, by typing information into a setup text file.

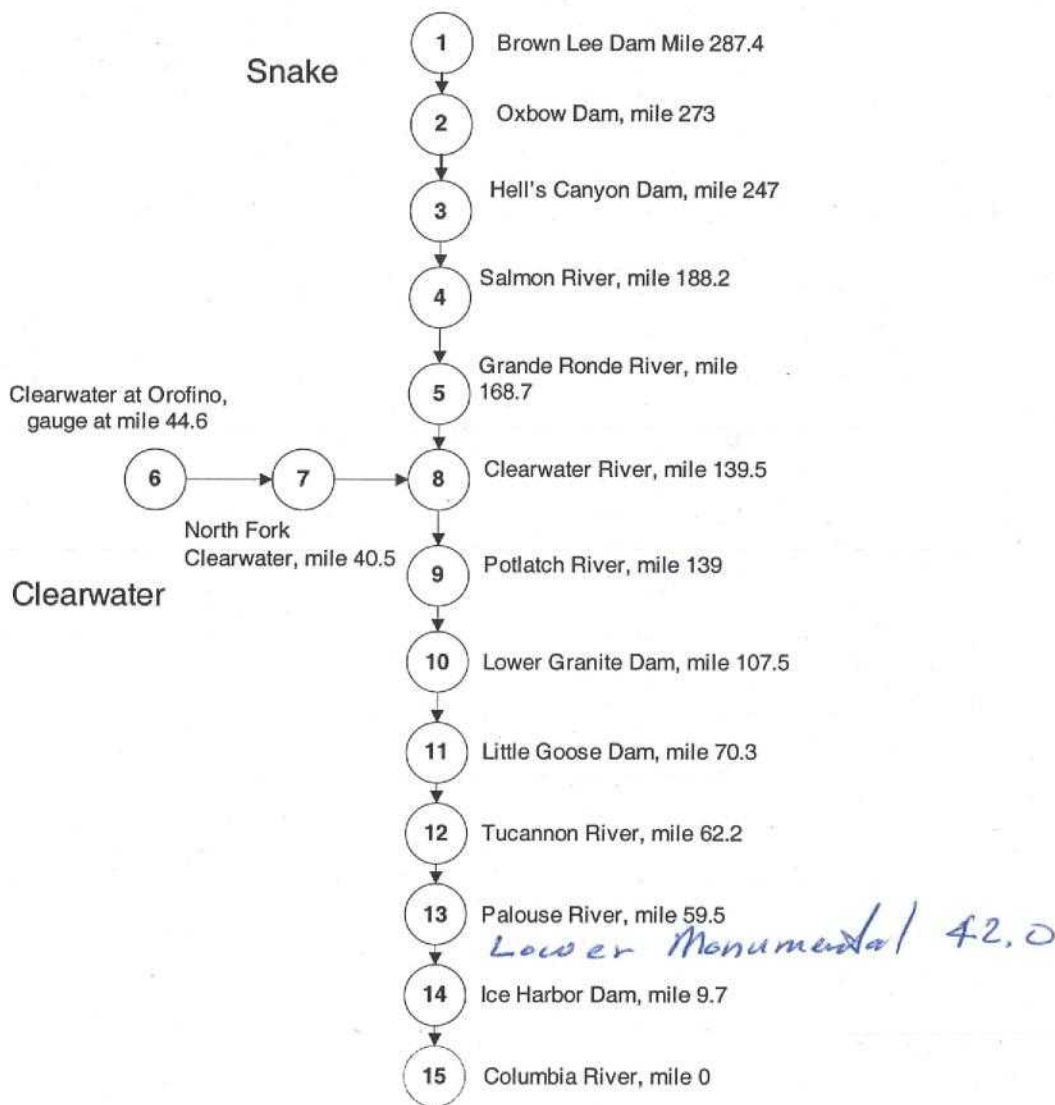
At its simplest level, creating a scenario involves the following steps:

- Choosing a river or rivers to simulate
- Choosing a time period over which to run the simulation
- Choosing boundary points on the river(s), i.e. a starting point and ending point
- Arranging tributary and dam nodes in a setup file (this amounts to a topological sort of the graph), so that data associated with tributaries and dams is consumed in the proper hierarchical order.

Such a scenario can be used to test the accuracy of the TMDL, by comparing predicted temperature values along a main stem, to actual recorded measurements. Of course, the real value of TMDL lies in its ability to make predictions based on “what if” scenarios such as removing a dam or dams.

The diagram on the following page represents a graph of the lower Snake River and Clearwater (not geographically accurate). Nodes represent data points where the program expects flow and temperature data (at tributaries), or elevation data (at dams). As indicated by the numbering, data is consumed by the TMDL program in topological order.

Scenario Visualization – Snake and Clearwater



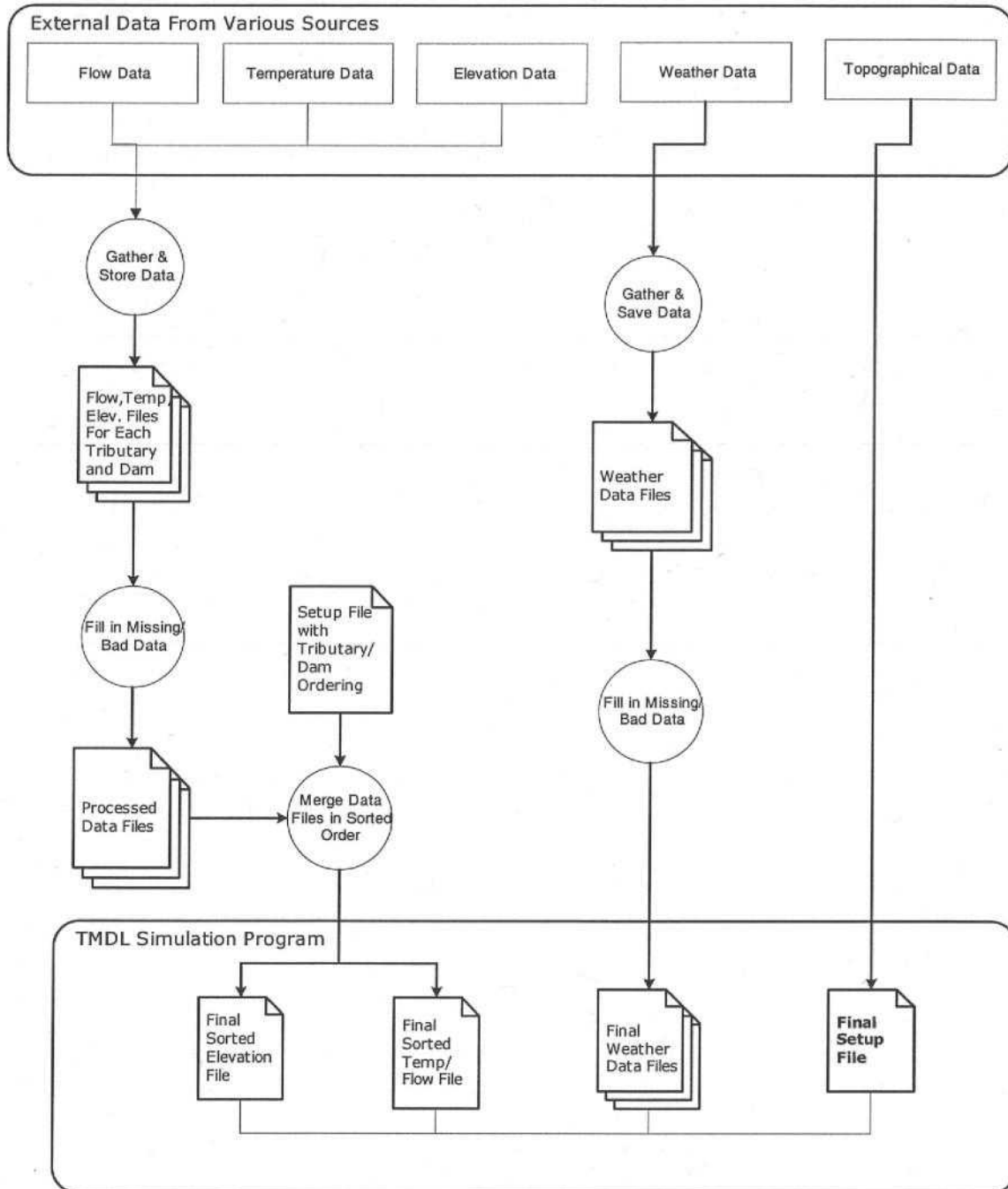
Data Sources

As one might expect, gathering the data necessary to run a TMDL temperature simulation is no small task. There exists no centralized database for river data. Rather, a handful of disparate state, federal and university agencies maintain Web sites that make available flow, temperature and elevation data taken from various river gauges at various dates and times. Such data is often incomplete or erroneous and is time intensive to collect.

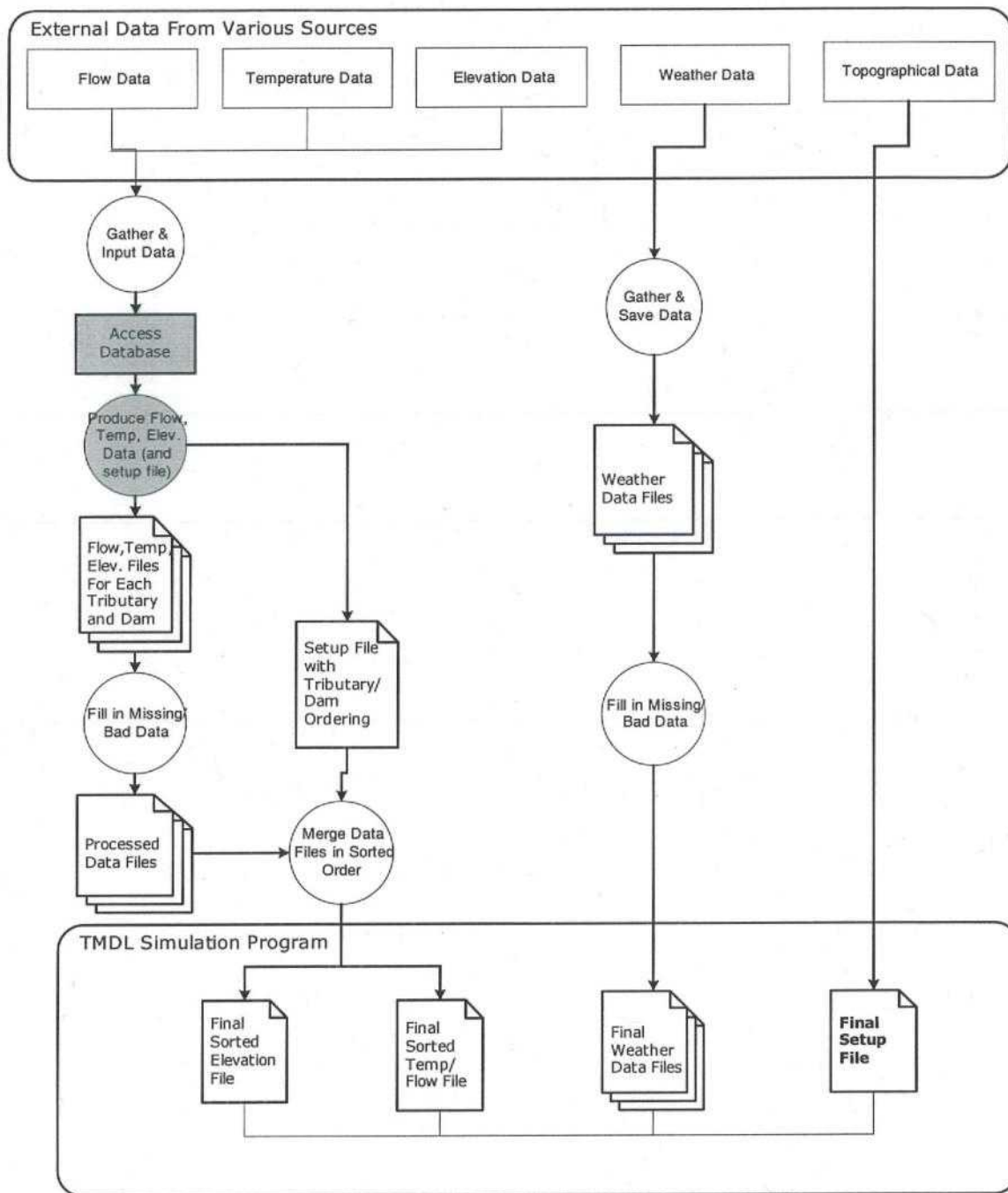
Once data is obtained, it must be processed to account for invalid or missing entries. Nonetheless, the TMDL team has assembled large amount data from past years. Currently, that data resides in flat text files.

Data Flow Diagram

The River Basin Model will initially provide flow, temperature and water elevation data to the TMDL. The following data flow diagram gives a top-level view of how temperature, flow and elevation data currently flow through the TMDL.



After implementation, the River Basin Model Service will provide functionality highlighted in yellow.



Summary

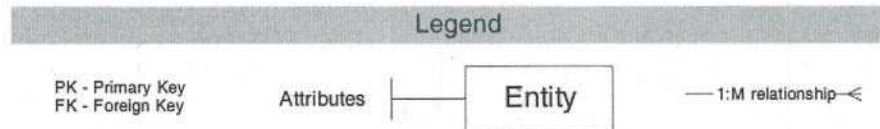
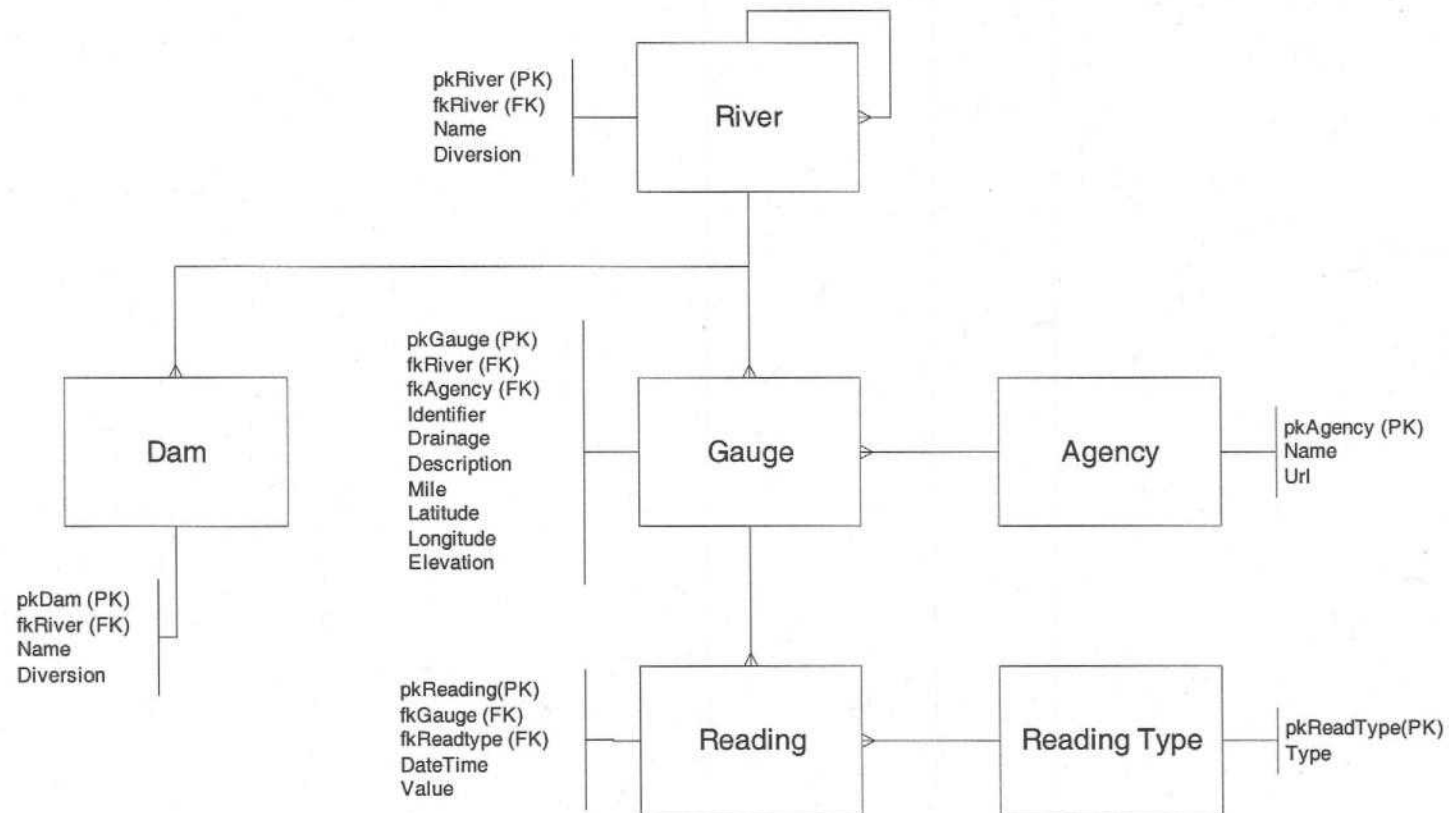
The River Basin Model Data Service will integrate into the existing TDML system. RBM will reduce data redundancy and data integrity problems by creating a centralized data

store that will contain all of the original flow, temperature and elevation data gathered from external sources. In addition, the River Basin Model Service will eliminate inefficiencies associated with finding and retrieving data from flat files. A graphical user interface will allow users to retrieve data necessary for running particular scenarios. Application logic will then write files that can be processed by existing utilities, fitting seamlessly into the existing system. Time permitting, application logic will also generate the initial setup file containing dams and tributaries in sorted order.

Appendix

See following pages for the RBM Entity Relationship Diagram and Logical Schema.

River Basin Model (ERD): Data Services for Columbia River Temperature TMDL



River Basin Logical Schema

